

Shoulder Instability

Interobserver and Intraobserver Agreement in the Assessment of Labral Tears

Brian R. Wolf,* MD, MS, Bastian Uribe, MD, Carolyn M. Hettrich, MD, MPH, Yubo Gao, PhD, Morgan Johnson, BA, and MOON Shoulder Group

Investigation performed at the University of Iowa, Iowa City, Iowa, USA

Background: The glenohumeral joint combines large range of motion and insufficient bony stabilization, making it susceptible to instability and dislocations. Arthroscopic surgery is routinely used as a diagnostic tool and has been considered the gold standard for the diagnosis of shoulder lesions. However, several studies have demonstrated variability in intraobserver and interobserver agreement.

Purpose: To evaluate interobserver and intraobserver agreement in the assessment of intra-articular lesions associated with shoulder instability among fellowship-trained shoulder surgeons.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: A total of 24 arthroscopic videos from patients treated for recurrent shoulder instability were shown to a group of 10 fellowship-trained shoulder surgeons who are members of the Multicenter Orthopaedic Outcomes Network (MOON) Shoulder Group. They were presented to the surgeons on 2 different occasions at least 2 months apart. They were asked to classify labral tears by their position, type, extension, other intra-articular abnormality, and preferred treatment. No patient history or physical examination data were provided. The primary outcome was the median overall percentage of agreement for the surgeons performing a video review, measured for each variable evaluated. Intraclass correlation coefficients were used to evaluate continuous variables, and kappa values were used for categorical items.

Results: Interobserver agreement was good for anterior labral lesions; good for Hill-Sachs lesions; and moderate for lesions of the superior labrum, posterior labrum, anterior sublabral foramen, and position and extension of the tear. Intraobserver agreement was either good or very good for all variables evaluated, except for being poor for inferior labral lesions and moderate for lesions of the meniscoid superior labrum.

Conclusion: Interobserver and intraobserver reliability for the arthroscopic assessment of labral tears in patients with recurrent shoulder instability were good to moderate for the majority of anatomic structures assessed. There was relatively good agreement between shoulder instability surgeons on assessing and documenting shoulder instability-associated abnormalities. These findings are important when interpreting collaborative clinical cohort studies with numerous surgeons involved in the research.

Keywords: shoulder; labrum; instability; agreement; reliability

The glenohumeral joint is the most commonly dislocated joint in the body, with an overall incidence of up to 23.9 per 100,000 per year. It is especially common among the young active male population (78%), with high redislocation rates after an initial traumatic dislocation.¹⁸

The glenohumeral joint combines large range of motion with insufficient bony stabilization, making it susceptible to instability and dislocations. It is stabilized by both static and dynamic mechanisms. The static mechanisms include the bony configurations of the glenoid and the humerus, the

glenoid labrum, the joint capsule, and the glenohumeral ligaments. The dynamic mechanisms include the muscles of the rotator cuff and, to a lesser degree, the long head of the biceps and the deltoid muscle.

A glenohumeral dislocation usually results in labral damage and may result in a fracture of the anteroinferior portion of the glenoid rim in association with soft tissue damage. This damage can contribute to recurrent anterior instability of the shoulder. Lesions of the glenoid rim and the corresponding Hill-Sachs impaction fracture of the humeral head have been reported in up to 97% and 90%, respectively, of initial dislocations in young athletes.¹⁵ For optimal outcomes, the surgeon must be well aware of the normal anatomy, the various anatomic variants, and the abnormality that can be encountered.¹⁶

The Orthopaedic Journal of Sports Medicine, 6(9), 2325967118793372
DOI: 10.1177/2325967118793372
© The Author(s) 2018

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

Arthroscopic surgery is routinely used as a diagnostic tool and has been considered the gold standard for the diagnosis of shoulder lesions in previous studies evaluating the diagnostic ability of different imaging techniques.^{5,6,8-10,12,13} However, studies evaluating shoulder arthroscopic surgery as a diagnostic tool have demonstrated variability in intraobserver and interobserver agreement when surgeons are describing and characterizing intra-articular structures.^{4,11,14,17} To our knowledge, only 1 study has evaluated this specifically for patients with shoulder instability, and that study only evaluated interobserver variability,¹⁴ while 2 other studies have evaluated patients specifically related to superior labrum anterior posterior (SLAP) lesions.^{4,17}

It is important to develop an understanding of interobserver reliability for clinical care, research, educational, and communication purposes. Interobserver agreement is relevant when orthopaedic surgeons interpret and apply the literature to their patients. More caution should be used in the assessment and treatment of lesions with low agreement. For multisurgeon collaborative research in which participant inclusion and exclusion may be related to intraoperative findings, an understanding of interobserver reliability and agreement among surgeons is crucial to ensuring that the same patients are being entered into the study or excluded across all participating sites.⁴ In addition, reliability data for collaborative multicenter cohort studies evaluating numerous intraoperative variables can provide insight into the validity of results determined from analysis of collected information. Videotaped surgical procedures are a useful tool for exploring interobserver reliability because they provide opportunity for assessing agreement.

The purpose of this study was to evaluate interobserver and intraobserver agreement among fellowship-trained shoulder surgeons in the assessment of intra-articular lesions, specifically labral lesions related to instability, in patients with shoulder instability. The goal of this study was to determine if these lesions could be identified and categorized reproducibly by surgeons performing arthroscopic surgery for shoulder instability.

METHODS

Twenty-four shoulder arthroscopic surgery videos corresponding to different patients with recurrent shoulder instability (anterior, posterior, and multidirectional) were compiled. Each video consisted of standard diagnostic arthroscopic surgery with visualization obtained through the posterior portal and the anterosuperior rotator interval portal. The videos demonstrated probing of all pertinent

anatomic structures, as done with routine shoulder diagnostic arthroscopic surgery. Diagnostic arthroscopic surgery was performed by 3 fellowship-trained surgeons (B.R.W., C.M.H., E.E.S.). There were 20 videos in the beach-chair position and 4 videos in the lateral decubitus position. Exemption from approval for this study was obtained from the University of Iowa Institutional Review Board.

The videos were shown to a group of 10 experienced sports medicine fellowship-trained shoulder surgeons who are members of the Multicenter Orthopaedic Outcomes Network (MOON) Shoulder Group. They were presented to the surgeons on 2 different occasions at least 2 months apart. For the second evaluation, the videos were arranged in a different order.

Surgeons were asked to classify labral tears by their type, position, extension, and preferred treatment based on the Shoulder Instability Operative Form used by the MOON Shoulder Group to record data in ongoing studies (Table 1). The position and extension of the tear had to be described on a visual representation of the glenoid: 360° divided evenly into 20 segments covering 18° each (Figure 1). There was no restriction to the number of diagnoses by structure that the surgeons could choose within the given alternatives in the questionnaire. There was no previously established “correct answer” for each of the videos. The reviewing surgeons performed their assessment independently and were blinded to the final treatment. No patient history or physical examination data were provided for the videos.

Two different basic modalities were used for the evaluation of the data obtained. As multiple options could be selected per variable, diagnostic categories were collapsed by anatomic region to calculate the kappa value for intraobserver and interobserver agreement (Table 2). For those items with a defined ordinal scale (ie, classification of SLAP tears), the original classification scheme was preserved.

Agreement for the type of lesion (frayed, cracked, detached, etc) within an anatomic region (anterior labrum, inferior labrum, etc) was expressed as a percentage of agreement, both within a surgeon’s own responses and between surgeons. Percentage of agreement between observers was chosen as the statistical representation of reliability because it provides a clear, straightforward, and easily interpreted statistical assessment of the data.³ The median pairwise agreement was selected over the mean as the primary measure to represent agreement for each variable because it is unaffected by extreme values.² Agreement among all possible surgeon pairwise combinations was calculated and interpreted with the scale previously described by Sasyniuk et al¹⁴ (Table 3). Statistical analysis was performed with SAS software (version 9.4; SAS Institute).

*Address correspondence to Brian R. Wolf, MD, MS, Department of Orthopedics and Rehabilitation, University of Iowa, 2701 Prairie Meadow Drive, Iowa City, IA 52242, USA (email: brian-wolf@uiowa.edu).

All authors are listed in the Authors section at the end of the article.

One or more of the authors has declared the following potential conflict of interest or source of funding: B.R.W. is a consultant for ConMed Linvatec, has received educational support from Wardlow Enterprises, and is on the Scientific Advisory Board for UnitedHealthcare. B.U. has received educational support from Arthrex. C.M.H. has received educational support from Arthrex, Zimmer Biomet, and Pacira Pharmaceuticals and has received research support from Tornier. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was waived by the University of Iowa Human Subjects Office/Institutional Review Board.

TABLE 1
Classification of Tear Types, Position/Extension, and Preferred Treatment Method^a

Diagnostic Categories	
Superior labrum (SLAP)	Normal, type I, type II, type III, type IV
Anterior labrum	Normal, frayed, cracked, detached, recessed (ALPSA), bucket handle, GLAD, absent
Inferior labrum	Normal, frayed, cracked, detached, recessed, bucket handle, absent
Posterior labrum	Normal, frayed, cracked, detached, recessed, bucket handle, absent
Hill-Sachs	No, yes (if present, size in mm)
Buford complex	No, yes
Anterior sublabral foramen	No, yes
Meniscoid superior labrum	No, yes
Position and extension of tear	360° divided evenly into 20 segments
Preferred treatment	Anterior arthroscopic Bankart repair: labrum only Anterior arthroscopic Bankart repair: labrum and capsule Posterior arthroscopic Bankart repair: labrum only Posterior arthroscopic Bankart repair: labrum and capsule Arthroscopic SLAP repair Arthroscopic debridement Arthroscopic suture plication Open Bristow/Latarjet procedure Open Bankart repair Open capsulorrhaphy Inferior capsular shift Open bone graft Other

^aALPSA, anterior labral periosteal sleeve avulsion; GLAD, glenolabral articular disruption; SLAP, superior labrum anterior posterior.

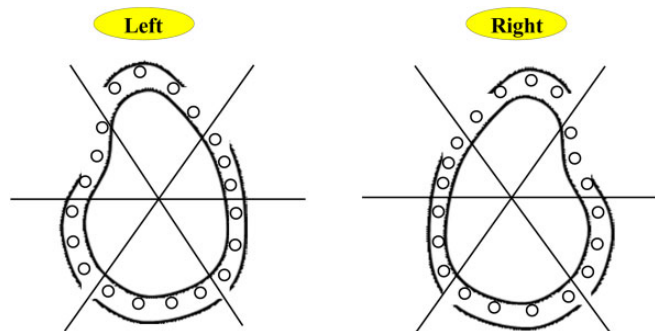


Figure 1. Diagrams used by surgeons to document location of labrum pathology.

TABLE 2
Categories of Intraobserver and Interobserver Agreement Based on Kappa Coefficient

Kappa	Agreement
<0.21	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

TABLE 3
Strength of Intraobserver and Interobserver Agreement Based on Value^a

Value of Agreement	Strength of Agreement
0%	None
1%-20%	Very poor
21%-40%	Poor
41%-60%	Good
61%-80%	Good
81%-99%	Very good
100%	Perfect

^aFrom Sasyniuk et al.¹⁴

TABLE 4
Intraobserver and Interobserver Agreement of Tear Type, Position/Extension, and Preferred Treatment Method^a

	Intrarater Kappa (95% CI)	Interrater Kappa (95% CI)			
Superior labrum (SLAP)	0.65 (0.52-0.79)	0.48 (0.46-0.50)			
Anterior labrum	0.84 (0.76-0.91)	0.76 (0.74-0.78)			
Inferior labrum	0.20 (0.09-0.31)	0.16 (0.15-0.18)			
Posterior labrum	0.58 (0.45-0.71)	0.50 (0.48-0.52)			
Hill-Sachs	0.79 (0.69-0.89)	0.66 (0.64-0.68)			
Buford complex	0.66 (0.53-0.79)	0.34 (0.32-0.36)			
Anterior sublabral foramen	0.63 (0.50-0.76)	0.48 (0.46-0.50)			
Meniscoid superior labrum	0.59 (0.53-0.64)	0.09 (0.07-0.11)			
Position and extension of tear	0.70 (0.65-0.76)	0.54 (0.54-0.55)			
Preferred treatment	0.67 (0.59-0.75)	0.40 (0.39-0.40)			
	Very Good	Good	Moderate	Fair	Poor

^aSLAP, superior labrum anterior posterior.

RESULTS

Intraobserver agreement was very good for anterior labral tears and good for the superior labrum, Hill-Sachs lesions, Buford complex, anterior sublabral foramen, position and extension of the tear, and preferred treatment (Table 4).

Intraobserver agreement was moderate for the posterior labrum and meniscoid superior labrum. Intraobserver agreement was poor for the assessment of the inferior labrum.

Interobserver agreement was good for the anterior labrum and Hill-Sachs lesions and moderate for the superior labrum, posterior labrum, anterior sublabral foramen, and position and extension of the tear. Interobserver agreement was fair for the Buford complex and poor for the inferior labrum and meniscoid superior labrum.

Agreement for the type of lesion (frayed, cracked, detached, etc) for each structure analyzed showed higher absolute values compared with the kappa coefficient. Median overall agreements were as follows: anterior labrum, 92%; inferior labrum, 67%; superior labrum, 75%; posterior labrum, 75%; and Hills-Sachs, 83%.

Overall agreement on preferred treatment was 67%. The largest difference observed for preferred treatment was variability in the inclusion of capsular plication along with labral repair. The second largest difference was moderate agreement on the diagnosis of SLAP tears, secondarily influencing treatment options. If these 2 factors are accounted for, overall agreement on preferred treatment increases to 75%.

DISCUSSION

Arthroscopic surgery is routinely used for the diagnosis of intra-articular abnormalities. Historical studies have demonstrated that routine magnetic resonance imaging misses labral lesions associated with shoulder instability in approximately 40% of cases; hence, arthroscopic surgery remains an important tool for diagnosing labral tears.^{1,7} In the current study, agreement among surgeons was the best for anterior labral lesions and Hill-Sachs lesions and worst for the inferior labrum. To our knowledge, this is the first study to present intraobserver reliability for labral lesions in a population with shoulder instability. Previous studies evaluating agreement for anterior, inferior, and posterior labral lesions have assumed higher intraobserver reliability of shoulder arthroscopic surgery as a diagnostic tool based on interobserver agreement, but a second round of evaluations was not performed.¹⁴ Intraobserver agreement was very good for the majority of variables evaluated; however, it was poor for the inferior labrum (kappa = 0.20) and moderate for the meniscoid superior labrum (kappa = 0.59) and posterior labrum (kappa = 0.58).

Sasyniuk et al¹⁴ showed varying interobserver agreement depending on the structure examined in the shoulder. They showed very good reliability (>80%) for the anterior labrum and Hill-Sachs lesions, poor agreement (<40%) for the glenoid and anteroinferior glenohumeral ligament, and intermediate for all other structures examined in a group of patients with anterior shoulder instability. Their findings indicated that the interpretation of intra-articular lesions in the glenohumeral joint could vary depending on the observer.¹⁴

We also described agreement for position and extension of the tear on a visual representation of the glenoid: 360° divided evenly into 20 segments covering 18° each. We showed this documentation technique to have good intraobserver reliability and moderate interobserver reliability.

This could indicate that surgeons have better agreement when shown which part of the labrum is affected on a visual model than they do when classifying it as an anatomic area, especially in the case of inferior labral lesions.

A wide spectrum of intraobserver variability in the classification and treatment of SLAP lesions was also noted by Gobezie et al,⁴ demonstrating moderate intraobserver reliability in both the classification and the treatment of labral injuries. When individual classifications and treatments were analyzed, the kappa values ranged from poor to moderate. The Gobezie et al⁴ study designated a correct answer for each case and analyzed agreement in each case accordingly. Wolf et al¹⁷ found that interobserver variability among experienced arthroscopic specialists evaluating the superior labrum and SLAP tears using the Snyder classification was significant, and analysis of intraobserver variability showed only moderate agreement for both diagnosis and treatment (kappa = 0.54 and 0.45, respectively). They also found that interobserver agreement decreased with the introduction of clinical vignettes. The Wolf et al¹⁷ study evaluated agreement, and no "correct" answer was assumed for each case. The results of the current study were similar, with moderate intraobserver agreement (kappa = 0.48) among surgeons.

The strengths of this study are that the video recordings present the same information to each surgeon for assessment. The limitations include the inability of the reviewer to probe structures themselves to determine their patency and the lack of history and physical examination data. Structures were probed in the videos; however, not having tactile feedback could negatively influence the ability of the surgeon to properly identify the abnormality. The only information given to surgeons was that all videos were performed in patients being treated for shoulder instability of some variety. In the actual clinical setting, treatment is decided by a combination of history and physical examination findings, including examination under anesthesia, imaging, and intraoperative results. This study only presents an analysis of one aspect in this equation.

Videos were of varying lengths and quality. However, each selected video was carefully reviewed and deemed to appropriately demonstrate all necessary anatomic structures by the senior author (B.R.W.). While all surgeons were fellowship trained and board certified, the level of experience of each surgeon was not evaluated. Even though previous studies¹⁴ have failed to prove a difference between more experienced and less experienced surgeons, they still advocate for the inclusion of only expert-level surgeons for reliability studies. Last, some videos were recorded in the beach-chair position and some were recorded in the lateral decubitus position. It is possible that surgeons' decisions were affected by the orientation of the videos depending on their preferred positioning for shoulder arthroscopic surgery in usual practice. It is also possible that poor agreement on assessing the inferior labrum could be related to the majority of the videos being recorded in the beach-chair position, as visualization of this area is more challenging in this position.

This study compared intraobserver reliability for each surgeon involved. As expected, it was higher than interobserver reliability, but several variables showed only

moderate correlation, and the inferior labrum showed poor agreement. The data collection and documentation of intraoperative findings can pose a challenge in multicenter collaboration research. Understanding the difficulties and limitations in acquiring and recording data is important for the interpretation of subsequent collaborative study outcomes. Overall, this study demonstrated good to moderate agreement in the assessment and treatment of most anatomic structures related to shoulder instability. Some variability was demonstrated on treating labral tears, which is notable in terms of how much the results from this study are helpful in establishing agreement among enrolling surgeons and for establishing validity of future outcome studies from this cohort data set.

CONCLUSION

Interobserver and intraobserver reliability of intra-articular lesions associated with shoulder instability among fellowship-trained shoulder surgeons varied depending on the structure evaluated. Generally, agreement was good for the majority of structures evaluated.

AUTHORS

Brian R. Wolf, MD, MS (Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, Iowa, USA); Bastian Uribe, MD (Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, Iowa, USA); Carolyn M. Hettrich, MD, MPH (Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, Iowa, USA); Yubo Gao, PhD (Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, Iowa, USA); Morgan Johnson, BA (Department of Orthopedics and Rehabilitation, University of Iowa, Iowa City, Iowa, USA); and MOON Shoulder Group: John E. Kuhn, MD (Vanderbilt University, Nashville, Tennessee, USA); Charles L. Cox, MD (Vanderbilt University, Nashville, Tennessee, USA); Brian T. Feely, MD (University of California, San Francisco, San Francisco, California, USA); Julie Bishop, MD (The Ohio State University, Columbus, Ohio, USA); Grant Jones, MD (The Ohio State University, Columbus, Ohio, USA); Robert H. Brophy, MD (Washington University in St Louis, St Louis, Missouri, USA); Matthew V. Smith, MD (Washington University in St Louis, St Louis, Missouri, USA); Keith M. Baumgarten, MD (Orthopedic Institute, Sioux Falls, South Dakota, USA); and Edwin E. Spencer, MD (Knoxville Orthopaedic Clinic, Knoxville, Tennessee, USA).

REFERENCES

1. Chandnani VP, Yeager TD, DeBerardino T, et al. Glenoid labral tears: prospective evaluation with MRI imaging, MR arthrography, and CT arthrography. *AJR Am J Roentgenol.* 1993;161(6):1229-1235.
2. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol.* 1990;43(6):551-558.
3. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43(6):543-549.
4. Gobeze R, Zurakowski D, Lavery K, Millett PJ, Cole BJ, Warner JJ. Analysis of interobserver and intraobserver variability in the diagnosis and treatment of SLAP tears using the Snyder classification. *Am J Sports Med.* 2008;36(7):1373-1379.
5. Hunter JC, Blatz DJ, Escobedo EM. SLAP lesions of the glenoid labrum: CT arthrographic and arthroscopic correlation. *Radiology.* 1992;184(2):513-518.
6. Huylebroek J, van Hedent E, van Overschelde J. Correlation of computed arthrotomography with arthroscopy of the glenohumeral joint. *Acta Orthop Belg.* 1991;57 (suppl 1):83-88.
7. Imhoff AB, Hodler J. Correlation of MR imaging, CT arthrography, and arthroscopy of the shoulder. *Bull Hosp Jt Dis.* 1996;54(3):146-152.
8. Kirkley A, Litchfield R, Thain L, Spouge A. Agreement between magnetic resonance imaging and arthroscopic evaluation of the shoulder joint in primary anterior dislocation of the shoulder. *Clin J Sport Med.* 2003;13(3):148-151.
9. Kneisl JS, Sweeney HJ, Paige ML. Correlation of pathology observed in double contrast arthrotomography and arthroscopy of the shoulder. *Arthroscopy.* 1988;4(1):21-24.
10. Loehr SP, Pope TL Jr, Martin DF, et al. Three-dimensional MRI of the glenoid labrum. *Skeletal Radiol.* 1995;24(2):117-121.
11. Mohtadi NG, Jager FL, Sasyniuk TM, Hollinshead RM, Fick GH. The reliability between surgeons comparing arthroscopic and video evaluation of patients with shoulder impingement syndrome. *Arthroscopy.* 2004;20(10):1055-1062.
12. Mohtadi NG, Vellet AD, Clark ML, et al. A prospective, double-blind comparison of magnetic resonance imaging and arthroscopy in the evaluation of patients presenting with shoulder pain. *J Shoulder Elbow Surg.* 2004;13(3):258-265.
13. Quinn SF, Brown TF. Meniscal tears diagnosed with MR imaging versus arthroscopy: how reliable a standard is arthroscopy? *Radiology.* 1991;181(3):843-847.
14. Sasyniuk TM, Mohtadi NG, Hollinshead RM, Russell ML, Fick GH. The inter-rater reliability of shoulder arthroscopy. *Arthroscopy.* 2007;23(9):971-977.
15. Taylor DC, Arciero RA. Pathologic changes associated with shoulder dislocations: arthroscopic and physical examination findings in first-time, traumatic anterior dislocations. *Am J Sports Med.* 1997;25(3):306-311.
16. Tischer T, Vogt S, Kreuz PC, Imhoff AB. Arthroscopic anatomy, variants, and pathologic findings in shoulder instability. *Arthroscopy.* 2011;27(10):1434-1443.
17. Wolf BR, Britton CL, Vasconcellos DA, Spencer EE; MOON Shoulder Group. Agreement in the classification and treatment of the superior labrum. *Am J Sports Med.* 2011;39(12):2588-2594.
18. Zacchilli MA, Owens BD. Epidemiology of shoulder dislocations presenting to emergency departments in the United States. *J Bone Joint Surg Am.* 2010;92(3):542-549.